

1 TGGCTCAGTCT CTGCTCACTT TGGCTCAGTCT AATGAGTAACT TTGGGAGTTT  
 51 CTTCTTCTCTT CTGCTCACTT TGGGAGTCTT CACTGAGTAT TGTGAAATAT  
 101 CAGAGATGGA CAGCTATCTG GTAGAGAACT TGGGAGTCTT CTTCTTATCT  
 151 TGGATGGAGC GGTCTTCTCT GGACCACTTG AACCCTAGCA TCTATGTGGG  
 201 CCTAGGCTCT TCGATCTCTG AGGCTGGGAC CAGGAGAGAG CTCTACCTCT  
 251 ACAGGCTCAA GCTTGGTTAC CAGCAGTCTC TCTTACGCTC TGGCTTCAGC  
 301 CAGGATGAGC GTGATCTGCA GGGCAAGCTT TCGATCGGCT AGCTCGGCTT  
 351 CTAGCTGCTC GCTCTCAGAG CCAACIGTAA GTTCTCAGG GGGCACAAGG  
 401 GGGACAGGCT GGTCTCACAG CTCAAATCTT TCTTCCACGA TGAGAAGAGA  
 451 GCCATTGACA CAGCAGCCAT GGCAGGCTTG GCATTATCT GTCTGAAGCG  
 501 CTCAAACTTC AACCTCTGTC GGAGACAAGG GATCATCATG GCCATCAGAA  
 551 CAGTGCAGAG GGAGATCTTG AAGGCTCAGA CTTCTCAGG CCACTTTGGG  
 601 AAGTCTCAGA GCACTCTCTT GCACTTACAG TTCTCTAIGA CTTCCCCCAT  
 651 GCTTGGGCTA GAAGTCTCTT CAGGATGGA GCTTCTCTAT GATTTCCTAG  
 701 TGGGAGTCTT GAGGATGGA GCTTCTCTAT GATTTCCTAG GATTTCCTAG  
 751 TGGGAGTCTT TCTTGAACCA CAGGATGGA GCTTCTCTAT GATTTCCTAG  
 801 TCTTCTCTCT CAGGATGGA GCTTCTCTAT GATTTCCTAG GATTTCCTAG  
 851 AGAGCTCAGA GATCATCATG GCTTCTCTAT GATTTCCTAG GATTTCCTAG  
 901 CAGGATGGA GATCATCATG GCTTCTCTAT GATTTCCTAG GATTTCCTAG  
 951 CAGGATGGA GATCATCATG GCTTCTCTAT GATTTCCTAG GATTTCCTAG  
 1001 CAGGATGGA GATCATCATG GCTTCTCTAT GATTTCCTAG GATTTCCTAG  
 1051 AGGAGTCTCT GGCAGGCTTG GCTTCTCTAT GATTTCCTAG GATTTCCTAG  
 1101 TATGCTCTCT TACAGAGCCA AGGATGGA GATCATCATG GATGAGCTGG  
 1151 TTAGCTCTCT CCGCTGAGGC TCTTCTCTAT GATTTCCTAG GATTTCCTAG  
 1201 TATGCTCTCT CCGCTGAGGC TCTTCTCTAT GATTTCCTAG GATTTCCTAG  
 1251 CTTCTCTCTA CTTCTCTCTA CTTCTCTCTA CTTCTCTCTA CTTCTCTCTA  
 1301 AGGAGTCTCT CTTCTCTCTA CTTCTCTCTA CTTCTCTCTA CTTCTCTCTA  
 1351 AGGAGTCTCT CTTCTCTCTA CTTCTCTCTA CTTCTCTCTA CTTCTCTCTA  
 1401 GCTTCTCTCT CTTCTCTCTA CTTCTCTCTA CTTCTCTCTA CTTCTCTCTA  
 1451 CAGGATGGA GATCATCATG GCTTCTCTAT GATTTCCTAG GATTTCCTAG  
 1501 GAGGATGGA GATCATCATG GCTTCTCTAT GATTTCCTAG GATTTCCTAG  
 1551 GAGGATGGA GATCATCATG GCTTCTCTAT GATTTCCTAG GATTTCCTAG  
 1601 GCTTCTCTCT CTTCTCTCTA CTTCTCTCTA CTTCTCTCTA CTTCTCTCTA  
 1651 CAGGATGGA GATCATCATG GCTTCTCTAT GATTTCCTAG GATTTCCTAG  
 1701 AAAAAAAAAA AAAAAAAAAA AA  
 (SEQ ID NO:1)

# FEATURES:

5'UTP: 1 - 30  
 Start Codon: 31  
 Stop Codon: 1119  
 3'UTP: 1112

## Homologous proteins:

Top 10 BLAST Hits

	Score	E
CFA110800002453390 /altid=gi1127427 /def=ref X1_009922.2  tr...	752	0.0
CFA110800002453336 /altid=gi1293316 /def=gb AAE21526.1  transo...	732	0.0
CFA118000004926133 /altid=gi1339206 /def=gb AAA61057.1  L02648...	732	0.0
CFA1108000024042036 /altid=gi11255467 /def=gb AAH31175.1  AAH01...	731	0.0
CFA118000004926130 /altid=gi14507109 /def=ref INP_00346.1  tran...	727	0.0
CFA118000004926132 /altid=gi1339206 /def=gb AAA61056.1  L02647...	725	0.0
CFA118000005170902 /altid=gi17657639 /def=ref NP_05364.1  tran...	515	e-145
CFA118000005218941 /altid=gi14572434 /def=gb AAD23523.1  AF12128...	501	e-140
CFA1164000136745249 /altid=gi11368124 /def=ref NP_071979.1  tr...	481	e-134
CFA118000004926134 /altid=gi14507407 /def=ref INP_001053.1  tran...	108	2e-22

EST:

gi 10725490	/dataset=dbest /taxon=9606...	858	0.0
gi 10947399	/dataset=dbest /taxon=9606...	846	0.0
gi 9121897	/dataset=dbest /taxon=9606...	846	0.0
gi 13280819	/dataset=dbest /taxon=9606...	846	0.0
gi 13287907	/dataset=dbest /taxon=9606...	833	0.0
gi 13286505	/dataset=dbest /taxon=9606...	831	0.0
gi 8150776	/dataset=dbest /taxon=9606...	815	0.0
gi 5936410	/dataset=dbest /taxon=9606...	726	0.0
gi 6888875	/dataset=dbest /taxon=9606...	726	0.0
gi 6888872	/dataset=dbest /taxon=9606...	726	0.0

**EXPRESSION INFORMATION FOR MODULATORY USE:**

library source:

gi|10725490|adult adrenal gland  
gi|10947399|mammary gland  
gi|9121897|retinoblastoma  
gi|13280819|adenocarcinoma cell line  
gi|13287907|retinoblastoma  
gi|13286505|embryonal carcinoma, cell line  
gi|8150776|adult uterus  
gi|5936410|adult uterus  
gi|6888875|adult head\_neck  
gi|6888872|adult head\_neck

Tissue Expression:

Human leukocyte





1 MHHDAWPEEL LNVLGALTEM CEIPEDQGH VEPDQHLDP WMDRLSLPHL  
 51 NPTTYGRLR LSIQATTEET LYLHSLKLY QV LKSAFS EDDGQDQFE  
 101 SPMIALYLL ALRANCEPVR GHFGDRLVSQ LKWFLEDEER AIDTAAMASL  
 151 APTLERSNE MRCRPQITM AIRTVREEIL KACTPEGHEG NVYJTFALQ  
 201 EIMTAMRGA ELGTACLKAR VALLASLDQG AFQNALMISQ LLFVINHETY  
 251 LHLFFDCLA PRVMLEPAAE TIPQTQEIIS VTIQVLSLLP FYRQSISVLA  
 301 GATVEVIRE AHELGCETYE TQASLSGPHL TSMCKAAGE REFWQLRDP  
 351 IYTLQSIAD YRKDGETIE LRLVSW  
 (SEQ ID NO:3)

# **FEATURES:**

## **Functional domains and key regions:**

PDOC00005 PS00005 PKC\_PHOSPH\_SITE

Protein kinase C phosphorylation site

Number of matches: 2

1	75-77	SLR
2	174-176	TVR

PDOC00006 PS00006 CK2\_PHOSPH\_SITE

Casein kinase II phosphorylation site

Number of matches: 6

1	67-70	TKFD
2	90-93	SELD
3	174-177	TVEE
4	226-228	SLSD
5	249-252	TALD
6	302-305	STVE

PDOC00008 PS00008 MYFISTYL

N-myristoylation site

Number of matches: 7

1	12-13	SVLSAL
2	57-60	GLFLSS
3	86-91	GLAFSE
4	149-154	GLAFTC
5	193-198	SHVYST
6	209-214	GAELGT
7	230-235	GAFLNA

PDOC00009 PS00009 AMIDATION

Amidation site

162-163	PGRR
---------	------

## **SignalP results:**

Measure	Position	Value	Cutoff	Conclusion
max. C	19	0.602	0.37	YES
max. Y	19	0.702	0.34	YES
max. S	5	0.974	0.38	YES
mean S	1-18	0.949	0.46	YES

Most likely cleavage site between pos. 18 and 19: ALT-EM

## **BLAST Alignment to Top Hit:**

>CRA|108000024636233 /altid=g|1298316 /def=gb|AAE25526.1|  
 transcobalamin II, PC II [human, endothelial cells,  
 Peptide, 427 aa] .org=human /taxon=9606 /dataset=nraa  
 /length=427  
 Length = 427

Score = 732 bits (1870), Expect = 0.0

Identities = 376/427 (88%), Positives = 376/427 (88%), Gaps = 51/427 (11%)

Frame = +1

Query: 81 MEHLGAFLELLGVIGALTEMFEIPEMICHILVEFLAQHLLPWWTRKLEHINPTIYYVLEPL 210  
 Sbjct: 1 MEHLGAFLELLGVIGALTEMFEIPEMICHILVEFLAQHLLPWWTRKLEHINPTIYYVLEPL 60

Query: 111 SLLQAITEEDLYLHSLKLGYYQQCLLGSAFSEDDGCGQGFPMGQALALYLLALRANCEFVR 390  
 Sbjct: 61 SLLQAITEEDLYLHSLKLGYYQQCLLGSAFSEDDGCGQGFPMGQALALYLLALRANCEFVR 120

Query: 381 GHFGDRLVSQLKWELEDEKRAI----- 456  
 Sbjct: 121 GHFGDRLVSQLKWELEDEKRAIGHDHKCHPHTSYYYQYGLGILALCLHQKRVHFSVVDKLL 180

Query: 457 -----DTAAMAGLAFTCLKFSNFNPGRRQRITMAIRTVREEILKAQTPEGHF 597  
 Sbjct: 181 YAVEPFHQGHHSVDTAAMAGLAFTCLKFSNFNPGRRQRITMAIRTVREEILKAQTPEGHF 240

Query: 598 GNVYSTPLALQFLMTSPMRGAELGTACLKARVALLASLQDGAFQNALMISQLLPVLNHKT 777  
 Sbjct: 241 GNVYSTPLALQFLMTSPMRGAELGTACLKARVALLASLQDGAFQNALMISQLLPVLNHKT 300

Query: 778 YIDLIFPDCLAPVMLEPAAETIPQTQEIISVTLQVLSLLPPYFQSIISVLGASTVEDVLK 957  
 Sbjct: 301 YIDLIFPDCLAPVMLEPAAETIPQTQEIISVTLQVLSLLPPYFQSIISVLGASTVEDVLK 360

Query: 958 KAHELGGFTYETQASLSGPLYLTSVMGKAAGEEFWQLLEDNTELLQGIADYRPKDGETI 1137  
 Sbjct: 361 KAHELGGFTYETQASLSGPLYLTSVMGKAAGEEFWQLLEDNTELLQGIADYRPKDGETI 420

Query: 1138 ELRLVSW 1158  
 ELRLVSW  
 Sbjct: 421 ELRLVSW 427  
 (SEQ ID NO:6)

#### HMM results:

Model	Description	Score	E-value	N
PF01122	Eukaryotic cobalamin-binding protein	829.9	8.6e-246	2
CE00052	CE00052 lymphocyte_transmembrane_protein_KAP	3.2	2.9	1

#### Parsed for domains:

Model	Domain	seq-f	seq-t	hmm-f	hmm-t	score	E-value
CE00052	1/1	1	11 [.	1	11 [.	3.2	2.9
PF01122	1/2	1	142 [.	1	143 [.	296.0	4.6e-85
PF01122	2/2	143	376 .]	197	450 .]	531.8	4.8e-156

1 MRHLGAELEL LGVLAALTEM CEIPENDSHL VEKLGQHLLP WNDPLSLHL  
51 NPSNYAUREL SSLQAGTFED LYHSLMLGY QQCLLGSAFS ELDDTCQHP  
101 SMGLALAYLL ALKANWHOHF GHPTSYYYQY GLGILALCLH QFPVHDSVVD  
151 KLLYAVEEPH QGHHSVITAA MAGLAFTCLK RSNENFGRRQ RITKAIRTVR  
201 EEIILKAQTPE GHFNNVYSTP LALQFLMTSP MRGAELGTAC LKARVALLAS  
251 LQDGAQNAL NISQLLEVLN HKTYIDLIFF DCLAPRVMLE PAAETIPQTQ  
301 EILSVTHQVL SLPPHYRQSI SVLAGSTVED VLKKAHLLCG FTYETQASLS  
351 GPVLTSMCK AAGEREFWQL LRDPNTPLLQ GIADYREKDG ETIELRLVSW  
(SBQ ID NO:4)

#### FEATURES:

##### Functional domains and key regions:

PD000005 PS00005 PKC\_PHOSPHO\_SITE  
Protein kinase C phosphorylation site  
198-200 TVF

PD000006 PS00006 CK2\_PHOSPHO\_SITE  
Casein kinase II phosphorylation site  
Number of matches: 7

1	67-73	THFD
2	90-93	SEID
3	147-153	SVND
4	198-201	TVFE
5	250-253	SLGD
6	273-276	TYID
7	326-329	STVE

PD000008 PS00008 MYRISTYL  
N-myristoylation site  
Number of matches: 7

1	12-17	GVLCAL
2	57-62	GLFLSS
3	86-91	GSAFSE
4	173-178	GLAFTC
5	214-219	GNVYST
6	233-238	GAELGT
7	254-259	GAFQNA

PD000009 PS00009 AMIDATION  
Amidation site  
186-189 PGFR

PD000048 PS00468 COBALAMIN\_BINDING  
Eukaryotic cobalamin-binding proteins signature  
165-173 SVITAAMAGLAFTC

#### SignalP results:

Measure	Position	Value	Cutoff	Conclusion
max. C	19	0.602	0.37	YES
max. Y	19	0.702	0.34	YES
max. S	5	0.974	0.88	YES
mean S	1-18	0.949	0.48	YES

Most likely cleavage site between pos. 13 and 19: ALT-EM

#### BLAST Alignment to Top Hit:

>CFA|108000024636236 /altid=gi|298316 /def=gb|AAB25526.1|  
transcobalamin II, TC II [human, endothelial cells,  
Peptide, 427 aa] /org=human /taxon=9606 /dataset=nraa  
/length=427  
Length = 427

Score = 120.1 bits (2026), Expect = 0.0  
 Identities = 399/427 (93%), Positives = 399/427 (93%), Gaps = 21/427 (5%)

Query: 1 MRHLCALFLFLGVLGALTEMCEIPEMDSHLVEKLGQHLLPWMDRLSLEHLNPSIYVGLRL 60  
 Subject: 1 MRHLCALFLFLGVLGALTEMCEIPEMDSHLVEKLGQHLLPWMDRLSLEHLNPSIYVGLRL 60

Query: 61 SSLQAGTKEDLYLHSLMDFTQQCLLGSAFSEDDGDCQGGKPMGQLALYLLALRAN----- 115  
 Subject: 61 SSLQAGTKEDLYLHSLMDFTQQCLLGSAFSEDDGDCQGGKPMGQLALYLLALRANCEFVR 120

Query: 116 -----W-----HDHKGHPHTSYQYGLGILALCLHJKEVHDSVVDKLL 153  
 Subject: 121 SHKGDRLVSQLKWFLEDEKRAIGHDHKHPHTSYQYGLGILALCLHJKEVHDSVVDKLL 180

Query: 154 YAEPEFHQGHHSVDTAAMAGLAFTCLNPSNENPGRQRITHAIRTVEEELKAQTPEGHF 213  
 Subject: 181 YAEPEFHQGHHSVDTAAMAGLAFTCLNPSNENPGRQRITHAIRTVEEELKAQTPEGHF 240

Query: 214 GNVYSTPLALQFILMTSPMGAELGTACLKARVALIASLQDGAFCNALMISQLLEVLNHKT 273  
 Subject: 241 GNVYSTPLALQFILMTSPMGAELGTACLKARVALIASLQDGAFCNALMISQLLEVLNHKT 300

Query: 274 VIDLIFFICLAFVVMLEPAAETIPQTQEIISVTLQQLSLLIFYFQSIYVLAGSTVEDVLK 333  
 Subject: 301 VIDLIFFICLAFVVMLEPAAETIPQTQEIISVTLQQLSLLIFYFQSIYVLAGSTVEDVLK 360

Query: 334 KAEELGGFTYETQASLSGPYLTSMVGKAAGEREFWQLLRDPNTELLQGIADYRPKDGETI 393  
 Subject: 361 KAEELGGFTYETQASLSGPYLTSMVGKAAGEREFWQLLRDPNTELLQGIADYRPKDGETI 420

Query: 394 ELRLVSW 400  
 Subject: 421 ELRLVSW 427  
 SEQ ID NO: 1)

#### HMM results:

Model	Description	Score	E-value	N
PF01122	Eukaryotic cobalamin-binding protein	90.13	8.6e-269	2
CE00052	CE00052 lymphocyte_transmembrane_protein_KAP	3.2	2.9	1

#### Parsed for domains:

Model	Domain	seq-f	seq-t	hmm-f	hmm-t	score	E-value
CE00052	1/1	1	11 [.]	1	11 [.]	3.2	2.9
PF01122	1/2	1	115 [.]	1	115 [.]	241.3	1.4e-68
PF01122	2/2	117	400 [.]	145	450 [.]	660.5	8.7e-195









10751 GAGTGTGGGG AAGAAAGAG TAAATGGAGA TCGCTGGGGA TATGCTTTCT  
10752 GTGATAGGGA GAAAGAAAT GTTGGGATGG ATGGGATGGT GGTATGATGG  
10753 ATGGGATGGAT GGGGAGAT GG ATGGAGGAGA ACACATTTTG GTTAAATGTA  
10801 ATACCAACATG ATAAAGGGA ATAGGAGAT GATCGAGGCT TTCTCTGAGA  
10851 GAGGGGCTGTA AGGATGAT GTGATTTTGG AATTAAAGAT GGAGGAGAG  
10901 GCGAGGTGGA GTGATATAT GCTGTAATGG CAACACTTTC GGAGGCGAG  
10951 GCGGGTGGGT GAGGTGAGGT GAGGAGTTGG AGACGAGGCT GGTAAATATG  
11001 GCGAAACTGG ATGTATTAAG AATAGAAAA AGTAGCTGGG TCGCTGGCG  
11051 AGTGGCTGTA AGGATAGGTA AGGATATAG GAGGATGAGG CAAGAGAGATC  
11101 ACTTGAACCT GAGAGGTGGA GGTGATATG AGGCAAGATC ATGGAATGG  
11151 ACTGCAATCT GGGTATAGGA ATAGGATAT GTCTAAAAA AAAAAAATAA  
11201 AAATGGAGAA GAGGAGAGT GGAATATGTT GCTGCTGCTT ATGATCTATG  
11251 GAGGTGAGG AGGTGAGGTA GATGAGTGG CTGAGGCTAG GAGGTGAGGA  
11301 GAGGCTGGG CAACATGCTG AAACTCTGCT TTTACTAAAA GAGGAGAGAT  
11351 GAGGCTGAGG TGGTGGTGA GAGGATGAT GAGGATATG AGGAGGCTG  
11401 AGGCAAGAA ATGATCTGAA GGTGGGAGAG AGAGGCTGGA GTGAGGCTGAG  
11451 ATGCTGCTG AGGATCTGAG GGTGGGCTAG AGGCTGAGG TCTGCTGCTG  
11501 GAAAAAAGAA GAATGAGAG AGTGGGCTGA AGAGAGAGGTA GAGGAGAGAA  
11551 AGAGAGAGAG GTGCAAGAGG TTTGGGCTGA TTTGGGCTGA TTTGGGCTGA  
11601 AAGAGGCTG AGTGGGCTGA ATGATGCTGA GTGGAGAGTA GATGAGGCTGA  
11651 GAGGAGCTGA GGTGAGGCTG GTAGAGAGTA TGGGCTGCTG AAGAGGCTGA  
11701 GTTGGGCTGA GAGGCTGCTG GAGGCTGCTG TGGGCTGCTG GTGGGCTGA  
11751 GATGATGAG AGGAGAGAGG GAGGCTGCTG GGTGCTGCTG GTGGGCTGA  
11801 GATGATGAG GAGGCTGCTG GAGGCTGCTG GGTGCTGCTG GTGGGCTGA  
11851 GATGATGAG GAGGCTGCTG GAGGCTGCTG GGTGCTGCTG GTGGGCTGA  
11901 GATGATGAG GAGGCTGCTG GAGGCTGCTG GGTGCTGCTG GTGGGCTGA  
11951 GATGATGAG GAGGCTGCTG GAGGCTGCTG GGTGCTGCTG GTGGGCTGA  
12001 GATGATGAG GAGGCTGCTG GAGGCTGCTG GGTGCTGCTG GTGGGCTGA  
12051 GATGATGAG GAGGCTGCTG GAGGCTGCTG GGTGCTGCTG GTGGGCTGA  
12101 GATGATGAG GAGGCTGCTG GAGGCTGCTG GGTGCTGCTG GTGGGCTGA  
12151 GATGATGAG GAGGCTGCTG GAGGCTGCTG GGTGCTGCTG GTGGGCTGA  
12201 GATGATGAG GAGGCTGCTG GAGGCTGCTG GGTGCTGCTG GTGGGCTGA  
12251 GATGATGAG GAGGCTGCTG GAGGCTGCTG GGTGCTGCTG GTGGGCTGA  
12301 GATGATGAG GAGGCTGCTG GAGGCTGCTG GGTGCTGCTG GTGGGCTGA  
12351 GATGATGAG GAGGCTGCTG GAGGCTGCTG GGTGCTGCTG GTGGGCTGA  
12401 GATGATGAG GAGGCTGCTG GAGGCTGCTG GGTGCTGCTG GTGGGCTGA  
12451 GATGATGAG GAGGCTGCTG GAGGCTGCTG GGTGCTGCTG GTGGGCTGA  
12501 GATGATGAG GAGGCTGCTG GAGGCTGCTG GGTGCTGCTG GTGGGCTGA  
12551 GATGATGAG GAGGCTGCTG GAGGCTGCTG GGTGCTGCTG GTGGGCTGA  
12601 GATGATGAG GAGGCTGCTG GAGGCTGCTG GGTGCTGCTG GTGGGCTGA  
12651 GATGATGAG GAGGCTGCTG GAGGCTGCTG GGTGCTGCTG GTGGGCTGA  
12701 GATGATGAG GAGGCTGCTG GAGGCTGCTG GGTGCTGCTG GTGGGCTGA  
12751 GATGATGAG GAGGCTGCTG GAGGCTGCTG GGTGCTGCTG GTGGGCTGA  
12801 GATGATGAG GAGGCTGCTG GAGGCTGCTG GGTGCTGCTG GTGGGCTGA  
12851 GATGATGAG GAGGCTGCTG GAGGCTGCTG GGTGCTGCTG GTGGGCTGA  
12901 GATGATGAG GAGGCTGCTG GAGGCTGCTG GGTGCTGCTG GTGGGCTGA  
12951 GATGATGAG GAGGCTGCTG GAGGCTGCTG GGTGCTGCTG GTGGGCTGA  
13001 GATGATGAG GAGGCTGCTG GAGGCTGCTG GGTGCTGCTG GTGGGCTGA  
13051 GATGATGAG GAGGCTGCTG GAGGCTGCTG GGTGCTGCTG GTGGGCTGA  
13101 GATGATGAG GAGGCTGCTG GAGGCTGCTG GGTGCTGCTG GTGGGCTGA  
13151 GATGATGAG GAGGCTGCTG GAGGCTGCTG GGTGCTGCTG GTGGGCTGA  
13201 GATGATGAG GAGGCTGCTG GAGGCTGCTG GGTGCTGCTG GTGGGCTGA  
13251 GATGATGAG GAGGCTGCTG GAGGCTGCTG GGTGCTGCTG GTGGGCTGA  
13301 GATGATGAG GAGGCTGCTG GAGGCTGCTG GGTGCTGCTG GTGGGCTGA  
13351 GATGATGAG GAGGCTGCTG GAGGCTGCTG GGTGCTGCTG GTGGGCTGA  
13401 GATGATGAG GAGGCTGCTG GAGGCTGCTG GGTGCTGCTG GTGGGCTGA  
13451 GATGATGAG GAGGCTGCTG GAGGCTGCTG GGTGCTGCTG GTGGGCTGA  
13501 GATGATGAG GAGGCTGCTG GAGGCTGCTG GGTGCTGCTG GTGGGCTGA  
13551 GATGATGAG GAGGCTGCTG GAGGCTGCTG GGTGCTGCTG GTGGGCTGA  
13601 GATGATGAG GAGGCTGCTG GAGGCTGCTG GGTGCTGCTG GTGGGCTGA  
13651 GATGATGAG GAGGCTGCTG GAGGCTGCTG GGTGCTGCTG GTGGGCTGA  
13701 GATGATGAG GAGGCTGCTG GAGGCTGCTG GGTGCTGCTG GTGGGCTGA  
13751 GATGATGAG GAGGCTGCTG GAGGCTGCTG GGTGCTGCTG GTGGGCTGA  
13801 GATGATGAG GAGGCTGCTG GAGGCTGCTG GGTGCTGCTG GTGGGCTGA  
13851 GATGATGAG GAGGCTGCTG GAGGCTGCTG GGTGCTGCTG GTGGGCTGA  
13901 GATGATGAG GAGGCTGCTG GAGGCTGCTG GGTGCTGCTG GTGGGCTGA  
13951 GATGATGAG GAGGCTGCTG GAGGCTGCTG GGTGCTGCTG GTGGGCTGA  
14001 GATGATGAG GAGGCTGCTG GAGGCTGCTG GGTGCTGCTG GTGGGCTGA  
14051 GATGATGAG GAGGCTGCTG GAGGCTGCTG GGTGCTGCTG GTGGGCTGA  
14101 GATGATGAG GAGGCTGCTG GAGGCTGCTG GGTGCTGCTG GTGGGCTGA  
14151 GATGATGAG GAGGCTGCTG GAGGCTGCTG GGTGCTGCTG GTGGGCTGA

FIGURE 3, page 4 of 22





21301 AGAAGAGAT TTTAGAGAT AGCAAGCTGGA TTGAAAAGAA AATTCTTGA  
 21351 GAGCTGGGTA TTTAAAGGGG GCTTTTCAGT TGGGCTTGAAG GTGGGATTCG  
 21401 GAGATAGGA AGAAGGAGAG AGA AGT AAC TGAGAGGTTA AATAA TTGG  
 21451 CGAGAGTCAG AGCAGGATG GAT AGA AG CTGGGCTTGA AATCGAGGCG  
 21501 TGTCTGTCAT TGGGCTTCG TTTCTA TTA GTGAAAGCTG AATCTGGGA  
 21551 GACTTCTGCG TGGGCTGGT TTTCTA TTA GACTGAGCTG AGAA GTTT  
 21601 TGTCTGTCAT TGTCTGAGG GTATTA TTA GAGAGTGGG GTGGCTGTCA  
 21651 GGTGTTTCTA TTAGCAATG AATCATTC GTGAGGAAG GTGA AGAA  
 21701 CTGAGTTAG AATCTGAGG AGGCTTGA GAGGTTCGA GCTTCTAAG  
 21751 ATCAAGGCG AGGCTTTCG TTA AGA TCG TACCTCAGG GTGGAGGT  
 21801 TGGAGATGA GGTTCGAGG TCGATATTC TCGTAATTC GCTCTCTCT  
 21851 GCTCAATCG TTTCTAGG TATCTTAC TACAGAAGCA AGGATAGAG  
 21901 AAGATTAAG CTGAGG TCG TTAAGTCTA GCTCTGAGG TCGCTATTC  
 21951 GAGAGGCG GAGAGTTCG TATCTTAC GCTCTGAGG TCGCTATTC  
 22001 GAGAGGCG GAGAGTTCG TATCTTAC GCTCTGAGG TCGCTATTC  
 22051 ATCTCTCTG GAGAGTTCG TATCTTAC GCTCTGAGG TCGCTATTC  
 22101 TTTCTAAGG TATCTTAC GAGAGTTCG TATCTTAC GAGAGTTCG  
 22151 GAGAGTTCG TATCTTAC GAGAGTTCG TATCTTAC GAGAGTTCG  
 22201 GAGAGTTCG TATCTTAC GAGAGTTCG TATCTTAC GAGAGTTCG  
 22251 GAGAGTTCG TATCTTAC GAGAGTTCG TATCTTAC GAGAGTTCG  
 22301 GAGAGTTCG TATCTTAC GAGAGTTCG TATCTTAC GAGAGTTCG  
 22351 GAGAGTTCG TATCTTAC GAGAGTTCG TATCTTAC GAGAGTTCG  
 22401 GAGAGTTCG TATCTTAC GAGAGTTCG TATCTTAC GAGAGTTCG  
 22451 GAGAGTTCG TATCTTAC GAGAGTTCG TATCTTAC GAGAGTTCG  
 22501 GAGAGTTCG TATCTTAC GAGAGTTCG TATCTTAC GAGAGTTCG  
 22551 GAGAGTTCG TATCTTAC GAGAGTTCG TATCTTAC GAGAGTTCG  
 22601 GAGAGTTCG TATCTTAC GAGAGTTCG TATCTTAC GAGAGTTCG  
 22651 GAGAGTTCG TATCTTAC GAGAGTTCG TATCTTAC GAGAGTTCG  
 22701 GAGAGTTCG TATCTTAC GAGAGTTCG TATCTTAC GAGAGTTCG  
 22751 GAGAGTTCG TATCTTAC GAGAGTTCG TATCTTAC GAGAGTTCG  
 22801 GAGAGTTCG TATCTTAC GAGAGTTCG TATCTTAC GAGAGTTCG  
 22851 GAGAGTTCG TATCTTAC GAGAGTTCG TATCTTAC GAGAGTTCG  
 22901 GAGAGTTCG TATCTTAC GAGAGTTCG TATCTTAC GAGAGTTCG  
 22951 GAGAGTTCG TATCTTAC GAGAGTTCG TATCTTAC GAGAGTTCG  
 23001 GAGAGTTCG TATCTTAC GAGAGTTCG TATCTTAC GAGAGTTCG  
 23051 GAGAGTTCG TATCTTAC GAGAGTTCG TATCTTAC GAGAGTTCG  
 23101 GAGAGTTCG TATCTTAC GAGAGTTCG TATCTTAC GAGAGTTCG  
 23151 GAGAGTTCG TATCTTAC GAGAGTTCG TATCTTAC GAGAGTTCG  
 23201 GAGAGTTCG TATCTTAC GAGAGTTCG TATCTTAC GAGAGTTCG  
 23251 GAGAGTTCG TATCTTAC GAGAGTTCG TATCTTAC GAGAGTTCG  
 23301 GAGAGTTCG TATCTTAC GAGAGTTCG TATCTTAC GAGAGTTCG  
 23351 GAGAGTTCG TATCTTAC GAGAGTTCG TATCTTAC GAGAGTTCG  
 23401 GAGAGTTCG TATCTTAC GAGAGTTCG TATCTTAC GAGAGTTCG  
 23451 GAGAGTTCG TATCTTAC GAGAGTTCG TATCTTAC GAGAGTTCG  
 23501 GAGAGTTCG TATCTTAC GAGAGTTCG TATCTTAC GAGAGTTCG  
 23551 GAGAGTTCG TATCTTAC GAGAGTTCG TATCTTAC GAGAGTTCG  
 23601 GAGAGTTCG TATCTTAC GAGAGTTCG TATCTTAC GAGAGTTCG  
 23651 GAGAGTTCG TATCTTAC GAGAGTTCG TATCTTAC GAGAGTTCG  
 23701 GAGAGTTCG TATCTTAC GAGAGTTCG TATCTTAC GAGAGTTCG  
 23751 GAGAGTTCG TATCTTAC GAGAGTTCG TATCTTAC GAGAGTTCG  
 23801 GAGAGTTCG TATCTTAC GAGAGTTCG TATCTTAC GAGAGTTCG  
 23851 GAGAGTTCG TATCTTAC GAGAGTTCG TATCTTAC GAGAGTTCG  
 23901 GAGAGTTCG TATCTTAC GAGAGTTCG TATCTTAC GAGAGTTCG  
 23951 GAGAGTTCG TATCTTAC GAGAGTTCG TATCTTAC GAGAGTTCG  
 24001 GAGAGTTCG TATCTTAC GAGAGTTCG TATCTTAC GAGAGTTCG  
 24051 GAGAGTTCG TATCTTAC GAGAGTTCG TATCTTAC GAGAGTTCG  
 24101 GAGAGTTCG TATCTTAC GAGAGTTCG TATCTTAC GAGAGTTCG  
 24151 GAGAGTTCG TATCTTAC GAGAGTTCG TATCTTAC GAGAGTTCG  
 24201 GAGAGTTCG TATCTTAC GAGAGTTCG TATCTTAC GAGAGTTCG  
 24251 GAGAGTTCG TATCTTAC GAGAGTTCG TATCTTAC GAGAGTTCG  
 24301 GAGAGTTCG TATCTTAC GAGAGTTCG TATCTTAC GAGAGTTCG  
 24351 GAGAGTTCG TATCTTAC GAGAGTTCG TATCTTAC GAGAGTTCG  
 24401 GAGAGTTCG TATCTTAC GAGAGTTCG TATCTTAC GAGAGTTCG  
 24451 GAGAGTTCG TATCTTAC GAGAGTTCG TATCTTAC GAGAGTTCG  
 24501 GAGAGTTCG TATCTTAC GAGAGTTCG TATCTTAC GAGAGTTCG  
 24551 GAGAGTTCG TATCTTAC GAGAGTTCG TATCTTAC GAGAGTTCG  
 24601 GAGAGTTCG TATCTTAC GAGAGTTCG TATCTTAC GAGAGTTCG  
 24651 GAGAGTTCG TATCTTAC GAGAGTTCG TATCTTAC GAGAGTTCG  
 24701 GAGAGTTCG TATCTTAC GAGAGTTCG TATCTTAC GAGAGTTCG  
 24751 GAGAGTTCG TATCTTAC GAGAGTTCG TATCTTAC GAGAGTTCG  
 24801 GAGAGTTCG TATCTTAC GAGAGTTCG TATCTTAC GAGAGTTCG  
 24851 GAGAGTTCG TATCTTAC GAGAGTTCG TATCTTAC GAGAGTTCG  
 24901 GAGAGTTCG TATCTTAC GAGAGTTCG TATCTTAC GAGAGTTCG  
 24951 GAGAGTTCG TATCTTAC GAGAGTTCG TATCTTAC GAGAGTTCG  
 25001 GAGAGTTCG TATCTTAC GAGAGTTCG TATCTTAC GAGAGTTCG

FIGURE 3, page 7 of 22





1:101,301,1:

# FEATURES:

Exon: 3831-5034  
 Intron: 5034-5164  
 Exon: 5164-5762  
 Intron: 5762-5991  
 Exon: 5991-7241  
 Intron: 7241-10000  
 Exon: 10001-10173  
 Intron: 10174-10298  
 Exon: 10299-10485  
 Intron: 10486-12027  
 Exon: 12028-12193  
 Intron: 12194-25821  
 Exon: 25822-25939

# Allelic Variants (SNPs):

DNA Position	Major	Minor	Domain	Protein Position	Major	Minor
901	C	T	Beyond ORF(5')			
1781	C	T	Beyond ORF(5')			
1850	G	A	Beyond ORF(5')			
2829	A	G	Intron			
3030	G	A	Intron			
6671	G	A	Intron			
6945	-	A	Intron			
6952	A	T	Intron			
7457	G	A	Intron			
7890	T	A	Intron			
8069	T	C	Intron			
8551	C	T	Intron			
9169	G	C	Intron			
9562	C	T	Intron			
9582	G	T	Intron			
11493	G	A T	Intron			
11760	A	G	Intron			
13086	T	C	Intron			
13183	T	C	Intron			
21240	C	G	Intron			
21695	A	G	Intron			
21698	C	T	Intron			
21733	C	A	Intron			
21245	C	-	Intron			
27375	C	T	Intron			
23042	A	- T	Intron			
23344	T	C	Intron			
23873	A	-	Intron			
26764	G	T	Intron			
24909	T	C	Intron			
24945	G	A	Intron			
25092	C	T	Intron			
25426	T	G	Intron			
25510	C	T	Intron			
25664	C	T	Intron			
26165	A	-	Beyond ORF(3')			

# Context:

DNA

Position

921 TTGGAGATATTTTAAAGGTCATAGTGTCTTCACAAATTGAGCTGAAAGGGAAGTGTTAGGA  
 TGATCTTGCCCTAACCCCTCTCATCTCACACAGGAAGAAGTATTTTAAACTCGAGAGGTTAA  
 GTGACCTGGCCTAAAGTCACACAGCCACCACTAGTTAACTCGTATACATTGATTCTCCTGT  
 GGGGCTGGGCAGATGAGGAATCTTTTGTCTCTTCCCTGTTTGCAGAGATTTTTTTTGTAG  
 GTTACTTTCCSAGTTCTGGCAAGTACCCCTGCTTCTGGTAGCTTTGTGTCTCGATTCAAT  
 [C,T]  
 TCATTCTTTTTATTTTATTTTATTTTGTAGACAGGGTCTCACTTTGTACCCCAAGCTGGA  
 GTGCAGTGGTGTAACTCTTGGCTCACTGTAGCCTCCAGCTCTTGGGTTCAGGAGATCCTTC



$\sigma_{\text{max}} = 1.5 \times 10^{-2}$

7457

7830

8089

8551

9269

FIGURE 3, page 11 of 22



GCTA CAGG TGGTGG AC TATGCG AGCGTATTTT TGTATTTT TAGTAGAGG AGGGTTT  
 TATT ATGTTTGGTAA TGGTTTGTAAACTT CCAAT TGGGTCGTCTGTGGGGGT  
 GCTA AGTGGTGGT AAGGGGATGGG GTT TGGG GG TATGTATTTATTTAGGCA  
 AGGT TGTGTGTGT TCGGGGCAAGG AAGGGCAAT TTAGGTAA TGTAGGCT  
 CAAT TATCGAAGT CCGGAGT TACGATG ACTA TACA CATATAT TTTTAT

[illegible][illegible]

22375 ACAAACCCCTTCCGAGACCTATACATATGCTCCACCTTGGAGCAAGAGCGGAGACATATTTTC  
CTGGGGAAGCTCTCTTCTCCAAATCTGTGGCGAGCTCGGCGCTGCAGTCTCTTCAAGAAAGG  
CAGCCCATGTCTCTGATCTGGATGAAGCATCTGCACCTCTCTTGCCTGAAAGATCTAGTTCG  
AGGCCGCAGCTGTGTTGAAGACCACTCGTTCTGTGGTGGGGTCTGTCAGAAAGGCTCTGC





Isoform 1:

FEATURES:

Exon: 2137-2139  
 Intron: 2139-5270  
 Exon: 5271-5273  
 Intron: 5274-1032  
 Exon: 7033-7061  
 Intron: 7062-9143  
 Exon: 9144-9301  
 Intron: 9302-10310  
 Exon: 10103-10374  
 Intron: 10375-10399  
 Exon: 10400-10586  
 Intron: 10587-10619  
 Exon: 12129-12294  
 Intron: 12295-25922  
 Exon: 25923-26040

Allelic Variants (SNPs):

DNA				Protein		
Position	Major	Minor	Domain	Position	Major	Minor
1-2	C	T	Beyond ORF(5')			
1-42	C	T	Beyond ORF(5')			
1-1	G	A	Beyond ORF(5')			
2-10	A	G	Intron			
3-21	G	A	Intron			
6732	G	A	Intron			
7053	G	A	Intron			
7-31	T	A	Intron			
8190	T	C	Intron			
8-12	C	T	Intron			
8-70	G	C	Intron			
9403	C	T	Intron			
8-83	G	T	Intron			
10134	G	A T	Intron			
10261	A	G	Intron			
11187	T	C	Intron			
11484	T	C	Intron			
11541	C	G	Intron			
11796	A	G	Intron			
11799	C	T	Intron			
11804	C	A	Intron			
11946	C	-	Intron			
12476	C	T	Intron			
12143	A	- T	Intron			
11445	T	C	Intron			
11974	A	-	Intron			
14165	G	T	Intron			
14140	T	C	Intron			
15046	G	A	Intron			
15193	C	T	Intron			
15119	T	G	Intron			
15614	C	T	Intron			
15735	C	T	Intron			
16136	A	-	Beyond ORF(3')			

Context:

RNA

Position

1022 TTGGAGATATTTTAAAGGTCATACTGTCTTCACAAATTGAGCTGAAAGGGAAGTGTAGGA  
 TGATCTTGCCCTAAGCCTCTCATCTCACACAGGAAGAAGTATTTTAACTCGAGAGGTTAA  
 GTGACCTGGCCAAAGTCACACAGCCACCACTAGTTAACTCGTATACATTGATTCTCTGT  
 GGGGCTGGGCAGATGAGGAATCTTTTGTCTCTTCCCTGTTTGCAGAGATTTTGTGAG  
 GTTACTTTCCGAGTTCTGGCAAGTACCCCTGCTTCTGGTAGCTTGTGTCTCGATTCAT  
 [C,T]  
 TCATTCTTTTATTTTATTTTATTTTGTGAGACAGGCTCTCATTGTGACCCGAAGTGGG

FIGURE 3, page 16 of 22









23445 A G G A A A C T G G T T T A A A A A G G A A A A G G G A C T C A C C A A G G T C A C A A A T A G G C A G T A T G  
C T G T G G G A A C T T G G C T A G G G G A C A C A G A C C T G G C C T G G G G C A G C C T G G C A G T C C T T A  
C T A A A T A C T G A A A A T C A G G G G C T T C G A T G A T G T T A T A A T G G A T G C A G A G C C C C A A A

TTAACTGAGAGGCTGGGA GGTAGAAATAGGCTTTACTTCTGTGTTTTCATAGGCTAG  
GATFAGGGGATCA GGTAGGCTTCTTGAAGAGAGGAGGATCTGGCTTTCTATTGTT  
[T, C]

GGGCTCTTA GGGGCTCTGCTGCTGCTTTT TTTGAAGAGGGGGTTGAGGCTAGGAGTAC  
GCTATGCTGACGCGACCAAGGTTAGGAAGAGGTCCTCTCTGAGTGGGGGCTTTTAAAG  
AAGAGTCTATGAGGCTCTCTCTGAGATCTCTTGGAAAGAGATGGTGACATTTAAATTT  
TTGAGGCGGATATTAAATAGCTAGAGCAACATCTTAAAGATTTAGGCTCTGAGAAAT  
TAGGGGAAGTAAATTTAAAGATTTCTTTATATTGGGCTAGGCTGATGGCTTACCT

24994

GGTCTGAGAAGATAGGGGAAGTAAGCAATTAAAAATTTCTTTATATTGGGCTAGTCT  
AATGCTCTCACTCTCTAATAGAGGCTTTTGGAGGAGTAAATGATGAGCTGAGGCTCAAG  
TTCAAGATGAGCTCTGCTAGCATGAGAGAAATGCTATTTAAATTTAAAGATTTAAAG  
CTGAGGGGCTCTGATCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCT  
TGGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCT  
[A, C]

AAAAAAGAAAAATATTTCTGAGAAAGAGCTAGGCTCTGAGAAATTTAAAGGAGTAAAT  
GAGCTGAGTAAATGAGAAATGAGAAATGAGAAATGAGAAATGAGAAATGAGAAATGAGAAAT  
CTCTGCT  
CT  
CT  
CT

24825

ATAGCTGAGATTAAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT  
GATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT  
GCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT  
TGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT  
CT  
[G, C]

GAGGCT  
GAGCT  
AGGCT  
GCT  
CTAAATTTCT

25040

GAGCT  
CT  
CT  
CT  
ATGAGCT  
CT  
[T, C]

AAAGGCT  
CT  
CT  
CT  
CT  
CT

25046

GCT  
CT  
GAGCT  
GAGCT  
AGGCT  
[G, A]

CT  
CT  
CT  
CT  
CT  
CT

25193

ATCATCT  
CT  
GCT  
GCT  
GCT  
[C, T]

GCT  
GCT  
GCT  
GCT  
GCT  
GCT

25529

AGTGAGAGAACTCAAAATTGAAGGAGAGAAAGCTCAGGAGGAGGGGAGTGTGAGTTGAGGA  
GTCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT

FIGURE 3, page 21 of 22

CGCTGTCTCTCTGATCGAGCGTTTACGCTGTCTCTCGTTGTGACATTGTAACAGATTCTAGGTGA  
GACTAGGAGAGTTCGTGATTCTCATTTCTCCGATAACATGAGCCGCCCAAGAGAGGACAAATTT  
CTGAGTGTATAACATAATATGATGTCGCCCTTCTCTGAGGGCATGCTGTATATTTAGCTTCGGA  
GAGATGCAAGCTTGAGCAGGCGAGATCGGGGGTCTAGAGCGGAAAGAGGCTGGAAGTT

[illegible]

FIGURE 3, page 22 of 22